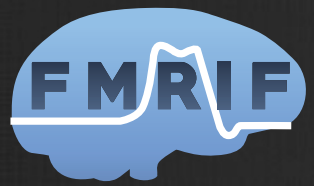
How Open Is My Science?

28 May 2024

Adam Thomas

Open Scientist in Residence Program, TOSI, The Neuro

Intramural Research Program
National Institute of Mental Health
Bethesda, Maryland, USA
http://cmn.nimh.nih.gov/dsst



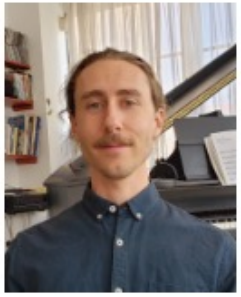
Thank You!

Tanenbaum Open Science Institute (TOSI)

Guy Rouleau
Director, The Neuro
Co-Founder of the Tanenbaum Open Science Institute

Annabel Seyller
Chief of Staff, The Neuro
CEO, Tanenbaum Open Science Institute

Jean-Baptiste Poline



Gabriel Pelletier
Open Science Data Manager
Interim Open Science Alliance Officer

Luisa Pimentel
Open Science Community Officer

NEURO SERIES
KILLAM SEMINARS

Lauréats KILLAM Laureates · neuro

Thank You!

Faculty

Jean-Baptiste Poline
Principal Investigator

Researchers

Brent McPherson
Postdoctoral researcher

Nikhil Bhagwat
Academic Associate

Staff

Sebastian Urchs
Research Software Developer

Alyssa Dai
Research Software Developer

Arman Jahanpour
Research Software Developer

Remi Gau
Research Associate

Graduate Students

Kendra Oudyk
PhD student

Ting Zhang
PhD student

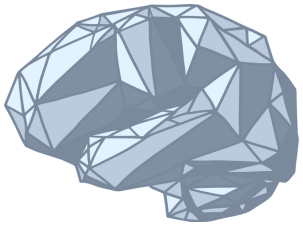
Jacob Sanz-Robinson
PhD student

Mohammad Torabi
PhD student

Emad Askarinejad
PhD student

Niusha Mirhakimi
MSc student

Michelle Wang
PhD student



ORIGAMI
Lab

My Team

Adam Thomas
Team Lead

Dustin Moraczewski
Data Scientist

Eric Earl
Data Scientist

Mia Zwally
Post-bac Trainee

Arshitha Basavaraj
Data Engineer

Jessica Dafflon
Data Scientist

?

YOU?

Interested in working with us? Reach out! http://bit.ly/adamt_osoh


The one-stop shop for
Open Science
in Neuroscience
OS
OH
Following

My Goal

Persuade you to use existing tools to evaluate openness and rigor in every project you're involved in

Slides & links: https://bit.ly/adamt_osoh

How Open is my Science?

Roadmap

- Why should you care?
- What evaluation tools are (and aren't) available?
- When should you use evaluation tools?
- What's on the horizon?

How Open is my Science? Roadmap

- Why should you care?
  - Open Science is important
  - Believing in it is not enough
- What tools are (and aren't) available?
- When should you use evaluation tool?
- What's on the horizon?

Open Science is important

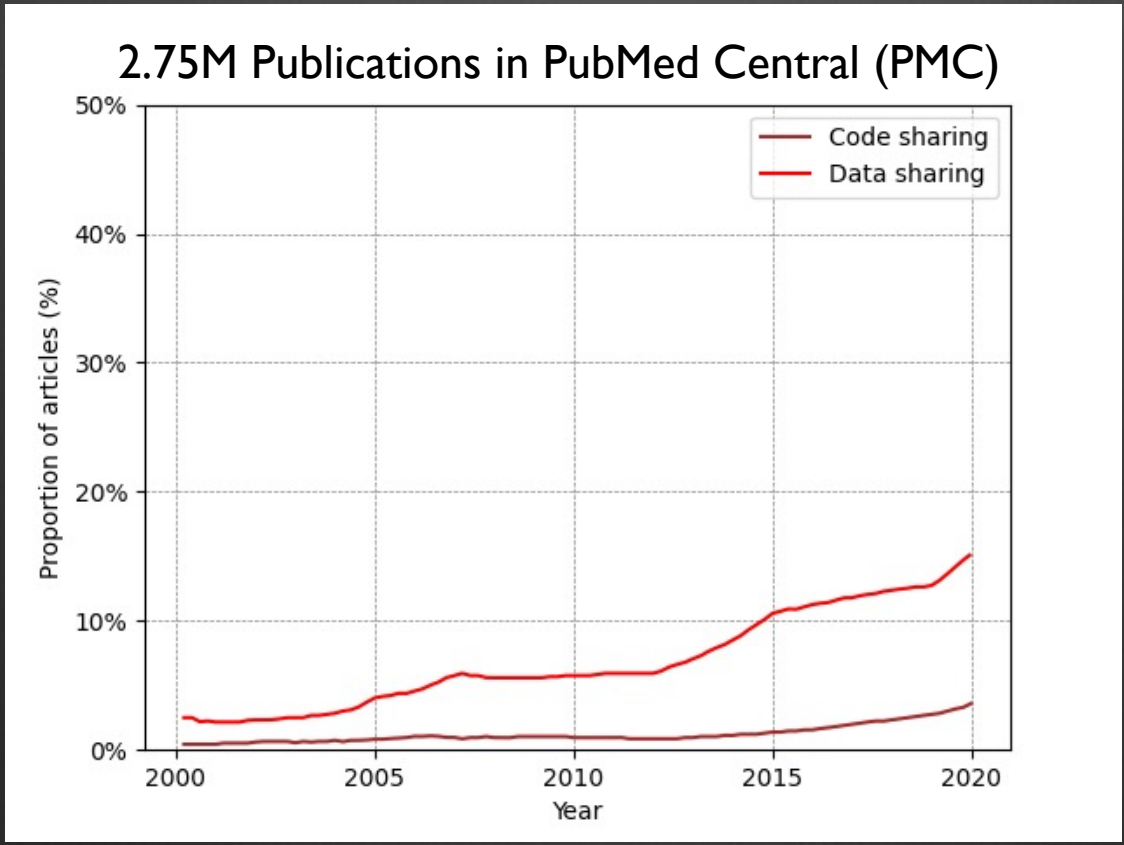
Data Sharing and Transparency in Policy and Practice at NIH

28 May 2024

Adam Thomas

Slides & links: https://bit.ly/adamt_osoh

Open Science is progressing, but slowly



Adapted from Serghiou et al. 2021 DOI:110.1371/journal.pbio.3001107

Believing is not enough
Open Science is complicated

https://en.wikipedia.org/wiki/Open_science

How Open is my Science? Roadmap

- Why should you care?
  - Open Science is important
  - Believing in it not enough
- What tools are (and aren't) available?
- What's on the horizon?

How Open is my Science? Roadmap

- Why should you care?
- What evaluation tools are (and aren't) available?
- When should you use evaluation tools?
- What's on the horizon?

Roadmap

- Why should you care?
- What tools are (and aren't) available?
- Why is "checking" useful and important?
- What's on the horizon?

Slides & links: http://bit.ly/adamt_osrp

Checklists

COBIDAS

(Committee on Best Practice in Data)

Best practices in data analysis and sharing in neuroimaging using MRI

Thomas E Nichols[1], Samir Das[2,3], Simon B Eickhoff[4,5], Alan C Evans[2,3], Tristan Glatard[2,6], Michael Hanke[7,8], Nikolaus Kriegeskorte[9], Michael P Milham[10,11], Russell A Poldrack[12], Jean-Baptiste Poline[13], Erika Proal[14], Bertrand Thirion[15], David C Van Essen[16], Tonya White[17] & B T Thomas Yeo[18]

NATURE NEUROSCIENCE VOLUME 20 | NUMBER 3 | MARCH 2017

Checklists
COBIDAS

Neuroimaging Methodological Reporting Checklist	Development Cycle						
	Problem Statement	Systematic Review	Steering Committee	Consensus Making	Pilot Testing	Adherence Monitoring	Updating Mechanism
COBIDAS-MRI	●	●	●	○	○	●	○
COBIDAS-MEEG	●	●	●	○	○	●	○
MRSinMRS	●	○	●	●	●	●	●
fMRI-FDCR	●	●	●	●	●	○	○
ContES	●	●	●	●	●	○	○
PET Checklist	●	○	●	○	●	●	●

Neuroimaging Methodological Reporting Checklist	Research Process Stages										
	Research Question	Research Design	Research Sample	Research Ethics	Measures/ Interventions	Data Collection	Data Analysis	Validating Findings	Results Reporting	Data Sharing	Reproducibility
COBIDAS-MRI	●	●	○	●	○	●	●	●	●	●	●
COBIDAS-MEEG	●	●	○	●	○	●	●	●	●	●	●
MRSinMRS	●	●	●	●	○	●	●	●	●	●	●
ContES	●	●	●	●	●	○	○	●	●	●	●
fMRI-FDCR	●	●	●	●	●	●	●	●	●	●	●
PET Checklist	●	●	●	●	●	●	●	●	●	●	●

Checklists

August 28, 1996

Improving the Quality of Reporting of Randomized Controlled Trials

The CONSORT Statement

Colin Begg, PhD; Mildred Cho, PhD; Susan Eastwood, ELS(D); et al

» Author Affiliations

JAMA. 1996;276(8):637-639. doi:10.1001/jama.1996.03540080059030

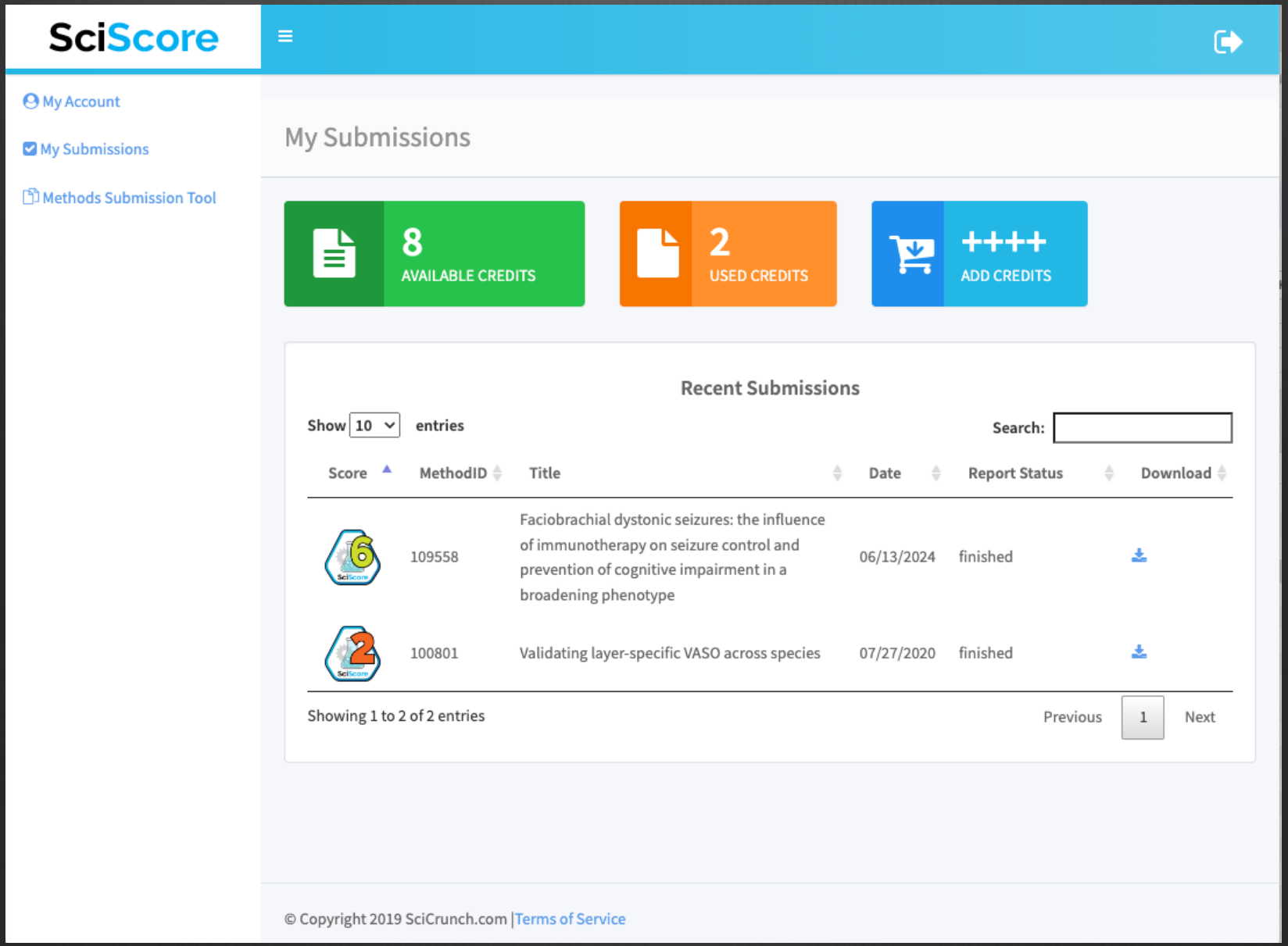
Checklists

- COBIDAS – Brain imaging
- CONSORT – Clinical trail
- MDAR
- STAR
- EQUATOR

SciScore Reports Supports

- SciScore Report
- CellPress STAR★Methods
- The MDAR (Materials Design Analysis Reporting) Framework for transparent reporting in the life sciences
- ARRIVE *In Vivo* animal Research (soon)

SciScore Reports



SciScore / MDAR Reports

SciScore: 6
What's this?

Document Identifier: **109558**

SciScore Report

Below you will find your SciScore report containing three tables. Your score is calculated based on adherence to scientific rigor criteria (Table 1) and identification of key biological resources (Table 2). Table 3 contains statistical tests and oligonucleotides but is not scored. If SciScore makes any mistakes, please contact us to help us learn and improve.

Table 1: Rigor Adherence Table

Ethics
IRB: Clinical data, serial Addenbrooke's cognitive examination revised (ACE-R) scores (Mioshi et al., 2006), videos documenting the events, results of investigations including detailed serological assays (methods in Vincent et al., 2004; Irani et al., 2010a), and treatment responses, were ascertained under Oxfordshire Research Ethics Committee A approval (07/Q160X/28).
Inclusion and Exclusion Criteria
not detected.
Attrition
For all but one patient, data were acquired using a 32-channel receive coil using a multi-echo MPRAGE sequence (van der Kouwe et al., 2008), with resolution 1 × 1 × 1 mm, repetition time = 2530 ms, inversion time = 1200 ms, echo time = 1.69, 3.55, 5.41 and 7.27 ms, matrix size = 256 × 256 × 176, and GRAPPA factor = 2.
Sex as a biological variable
not detected.
Subject Demographics
Age: Scans from an available sample of 13 healthy individuals (48 to 75 years of age) from multiple studies at the Dementia Research Centre were included for comparisons.
Randomization
not detected.
Blinding
not detected.
Power Analysis
not detected.
Replication
not required.

Design

Study protocol	Yes (indicate where provided: page no/section/legend)	n/a
For clinical trials, provide the trial registration number OR cite DOI in manuscript.	Not detected.	

Laboratory protocol	Yes (indicate where provided: page no/section/legend)	n/a
Provide DOI or other citation details if detailed step-by-step protocols are available.	Not detected.	

Experimental study design (statistics details)	Yes (indicate where provided: page no/section/legend)	n/a
State whether and how the following have been done, or if they were not carried out		
Sample size determination	not detected.	
Randomization	not detected.	
Blinding	not detected.	
inclusion/exclusion criteria	not detected.	

Sample definition and in-laboratory replication	Yes (indicate where provided: page no/section/legend)	n/a
State number of times the experiment was replicated in laboratory	Not detected.	
Define whether data describe technical or biological replicates	Not detected.	

Ethics	Yes (indicate where provided: page no/section/legend)	n/a
Studies involving human participants: State details of authority granting ethics approval (IRB or equivalent committee(s), provide reference number for approval.	Clinical data, serial Addenbrooke's cognitive examination revised (ACE-R) scores (Mioshi et al., 2006), videos documenting the events, results of investigations including detailed serological assays (methods in Vincent et al., 2004; Irani et al., 2010a), and treatment responses, were ascertained under Oxfordshire Research Ethics Committee A approval (07/Q160X/28).	
Studies involving experimental animals: State details of authority granting ethics approval (IRB or equivalent committee(s), provide reference number for approval.	Not detected.	
Studies involving specimen and field samples: State if relevant permits obtained, provide details of authority approving study; if none were required, explain why.	Not detected.	

Dual Use Research of Concern (DURC)	Yes (indicate where provided: page no/section/legend)	n/a
If study is subject to dual use research of concern, state the authority granting approval and reference number for the regulatory approval	Not currently checked by SciScore	

Roadmap

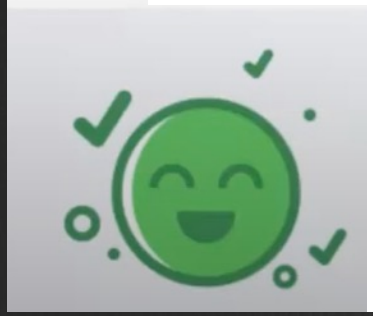
- Why should you care?
- What tools aren't available?
- When should you use these tools?
- What's on the horizon?

Lessons learned from automated screening

Lessons learned from automated screening

User feedback:

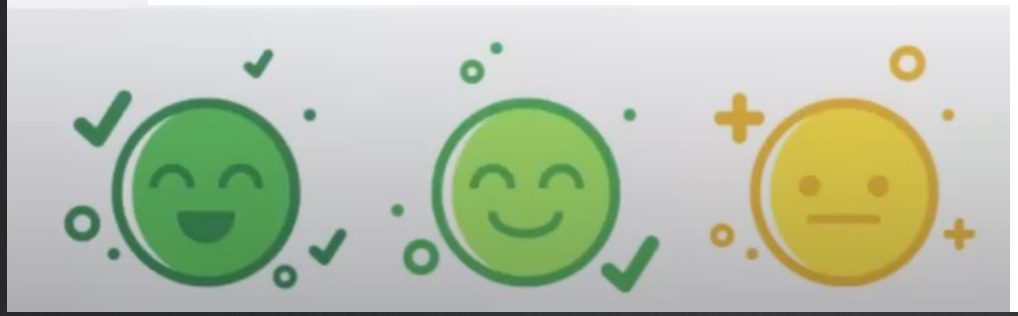
- Thanks for checking my paper



Lessons learned from automated screening

User feedback:

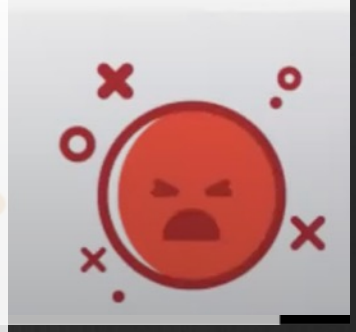
- Thanks for checking my paper
- You have missed something



Lessons learned from automated screening

User feedback:

- Thanks for checking my paper
- You have missed something
- I can't believe you dare to read my paper using some stupid bot!

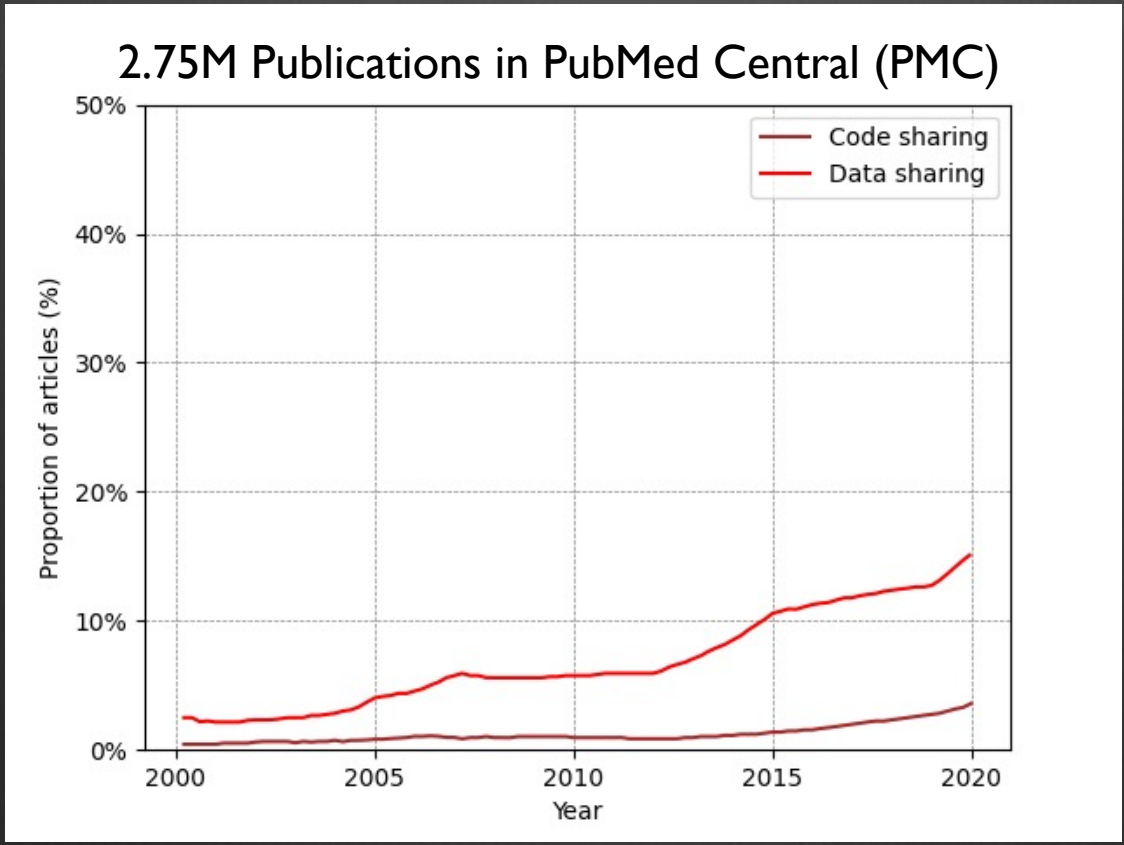


Lessons learned from automated screening

- Most authors are acting in good faith
- All automated tools will make errors
- Many authors will want to correct the manuscript or the tool's error
- Potential to create ill will

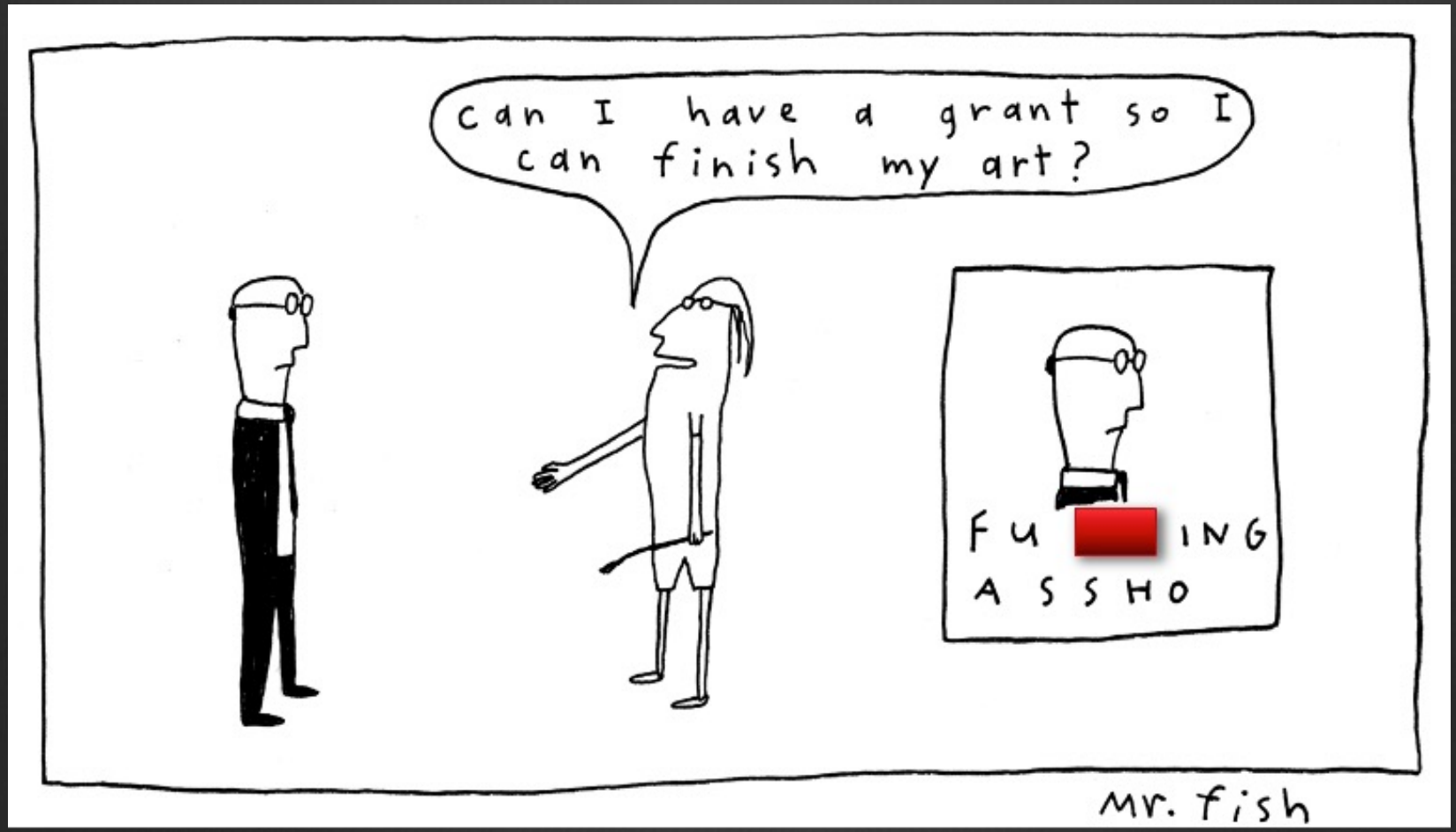
https://youtu.be/581NkRV4mbA

Open science shaming have unintended consequences



Adapted from Serghiou et al. 2021 DOI:110.1371/journal.pbio.3001107

Open science shaming have unintended consequences

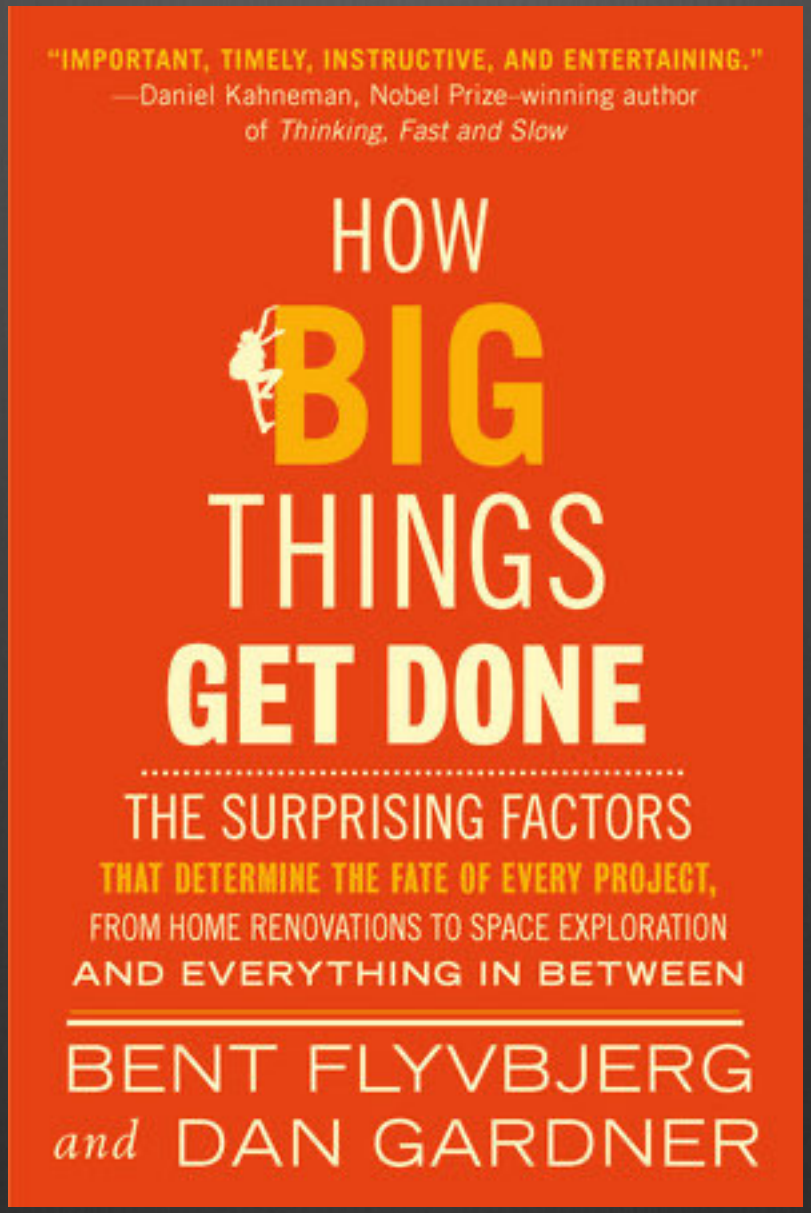


Support Mr. Fish! https://clowncrack.com/about/

How Open is my Science? Roadmap

- Why should you care?
- What evaluation tools are (and aren't) available?
- When should you use evaluation tools?
- What's on the horizon?

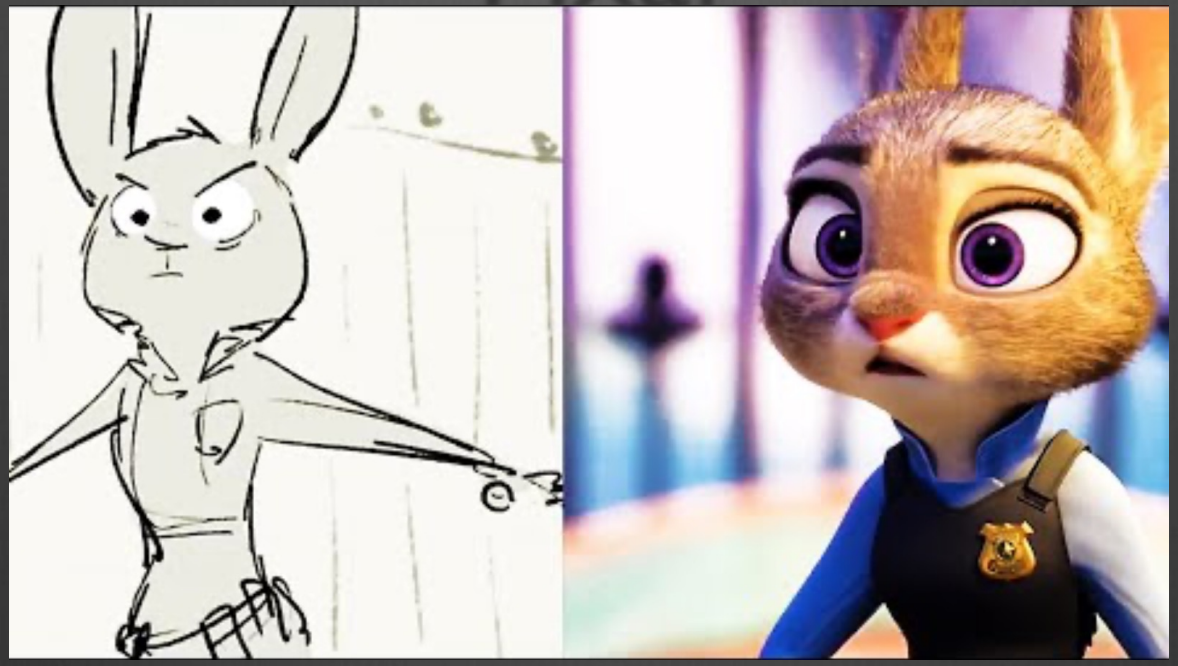
The importance of iteration



Pixar

Film	U.S. release date	Box office gross	Budget	% Profit
Toy Story	1995	$373 M	$30 M	1,245%
Toy Story 2	1999	$497 M	$90 M	553%
Toy Story 3	2010	$1,066 M	$200 M	533%
Toy Story 4	2019	$1,0733 M	$200 M	537%
Total		$ 3 B	$520 M	576%

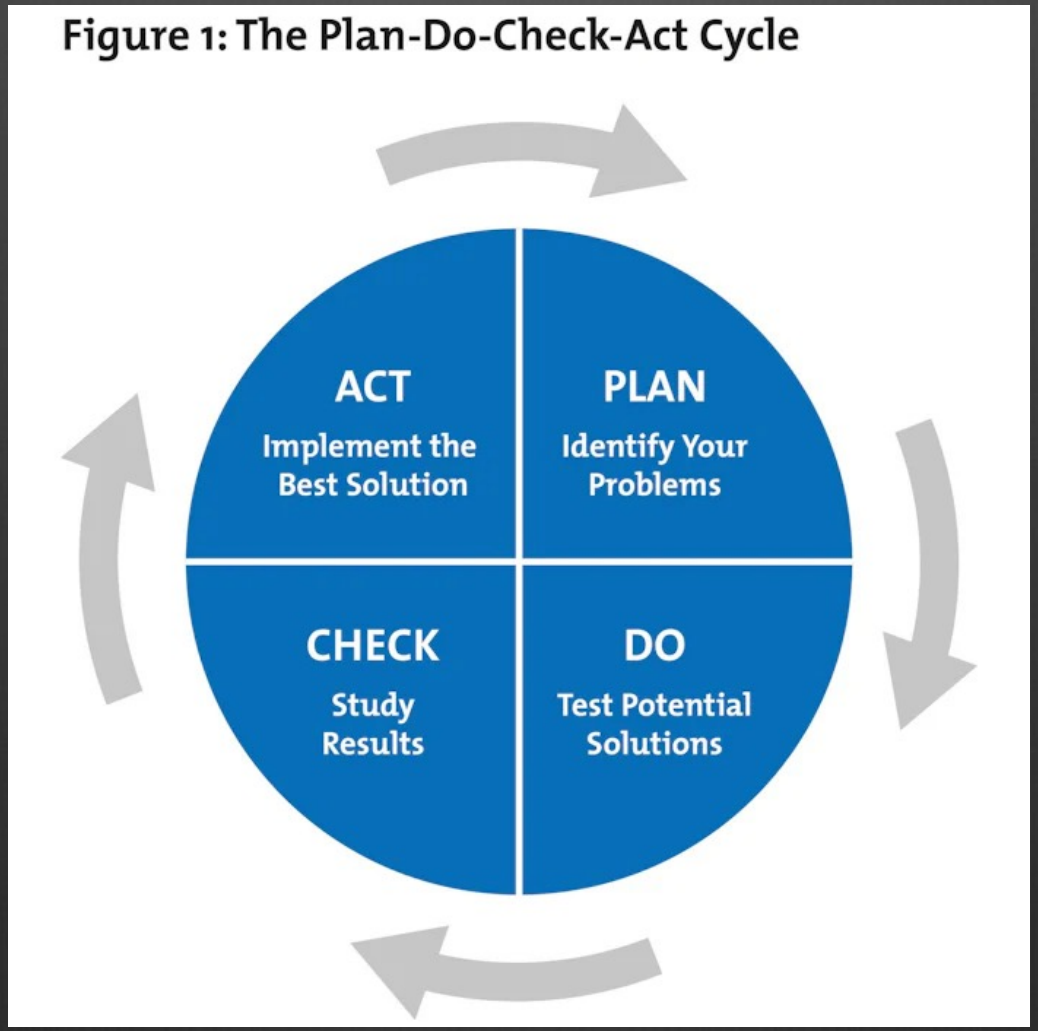
Pixar



https://youtu.be/_rTNT4J2FGo?si=YZQwKgDxcp3a-wOZ

- Write a script
- Storyboard version of the whole movie is created
  - Approximate 2,700 storyboards
  - Three months work for 6-8 people
  - Pixar staff watch the movie and give feedback
  - Process repeated an average of eight times

The Plan-Do-Check-Act Cycle



Evaluated before you start: Registered Reports

nature neuroscience

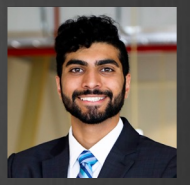
A positive-negative mode of population covariation links brain connectivity, demographics and behavior

Stephen M Smith[1], Thomas E Nichols[2], Diego Vidaurre[3], Anderson M Winkler[1], Timothy E J Behrens[1], Matthew F Glasser[4], Kamil Ugurbil[5], Deanna M Barch[4], David C Van Essen[4] & Karla L Miller[1]

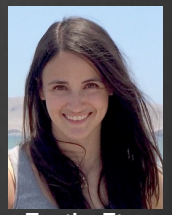
Computationally replicating the Smith et al. (2015) positive-negative mode linking functional connectivity and subject measures

bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

Nikhil Goyal[1], Dustin Moraczewski[1], Peter Bandettini[2], Emily S. Finn[2], Adam Thomas[1]
[1]Data Science and Sharing Team, National Institute of Mental Health, Bethesda, MD
[2]Section on Functional Imaging Methods, National Institute of Mental Health, Bethesda, MD


Nikhil Goyal


Dustin Moraczewski


Emily Finn


Peter Bandettini

ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Replication

Check for updates

Cite this article: Goyal N, Moraczewski D, Bandettini PA, Finn ES, Thomas AG. 2022 The positive–negative mode link between brain connectivity, demographics and behaviour: a pre-registered replication of Smith *et al.* (2015). *R. Soc. Open Sci.* **9**: 201090.
https://doi.org/10.1098/rsos.201090

The positive–negative mode link between brain connectivity, demographics and behaviour: a pre-registered replication of Smith *et al.* (2015)

Nikhil Goyal[1,†], Dustin Moraczewski[1,†], Peter A. Bandettini[2], Emily S. Finn[2,3] and Adam G. Thomas[1]

How Open is my Science?

Roadmap

- Why should you care?
- What evaluation tools are (and aren't) available?
- When should you use evaluation tools?
- What's on the horizon?

Large Language Models (LLM) offer dramatic improvements in content extraction

Many Open Science Rubrics require judgement

- FAIR (Wilksinson et al, 2016)

ID	Questions from https://www.go-fair.org/fair-principles/
F1	(Meta)data are assigned a globally unique and persistent identifier
F2	Data are described with rich metadata (defined by R1 below)
F3	Metadata clearly and explicitly include the identifier of the data they describe
F4	(Meta)data are registered or indexed in a searchable resource
A1	(Meta)data are retrievable by their identifier using a standardised communications protocol
A1-1	The protocol is open, free, and universally implementable
A1-2	The protocol allows for an authentication and authorisation procedure, where necessary
A2	Metadata are accessible, even when the data are no longer available
I1	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
I2	(Meta)data use vocabularies that follow FAIR principles
I3	(Meta)data include qualified references to other (meta)data
R1	(Meta)data are richly described with a plurality of accurate and relevant attributes
R1-1	(Meta)data are released with a clear and accessible data usage license
R1-2	(Meta)data are associated with detailed provenance
R1-3	(Meta)data meet domain-relevant community standards

LLMs can answer Nuanced Open Science Question Today

Link to Shared Data	Does the dataset include detailed provenance?	Does the metadata meet domain-relevant community standards?
- Online Participants' data: https://osf.io/km69z - Mobile App Participants' data: https://doi.org/10.5061/dryad.prr4xgxkk	Yes, the dataset includes detailed provenance as the data analysis software, data from participants, and code used in the study are all available on public platforms.	Yes, the metadata meets domain-relevant community standards as it provides detailed information on data and code availability, uses recognized repositories and platforms for sharing, and includes a clear data availability statement along with a declaration of no competing interests.

Validation requires Manual Labelling

Mining the neuroimaging literature.

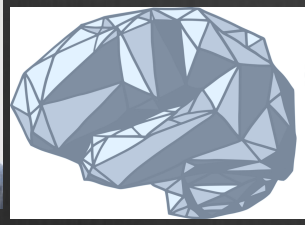
Jérôme Dockès[1*], Kendra Oudyk[2*], Mohammad Torabi[2], Alejandro I de la Vega[3], and Jean-Baptiste Poline[2]

[1]National Institute for Research in Digital Science and Technology (INRIA), Paris, France
[2]Montreal Neurological Institute, McGill University, Montreal, Canada
[3]University of Texas at Austin, Austin, Texas, USA
[*]co-first authors

pubget

build passing | codecov 100% | pubget on GitHub

labelbuddy



ORIGAMI Lab

Kendra Oudyk JB Poline

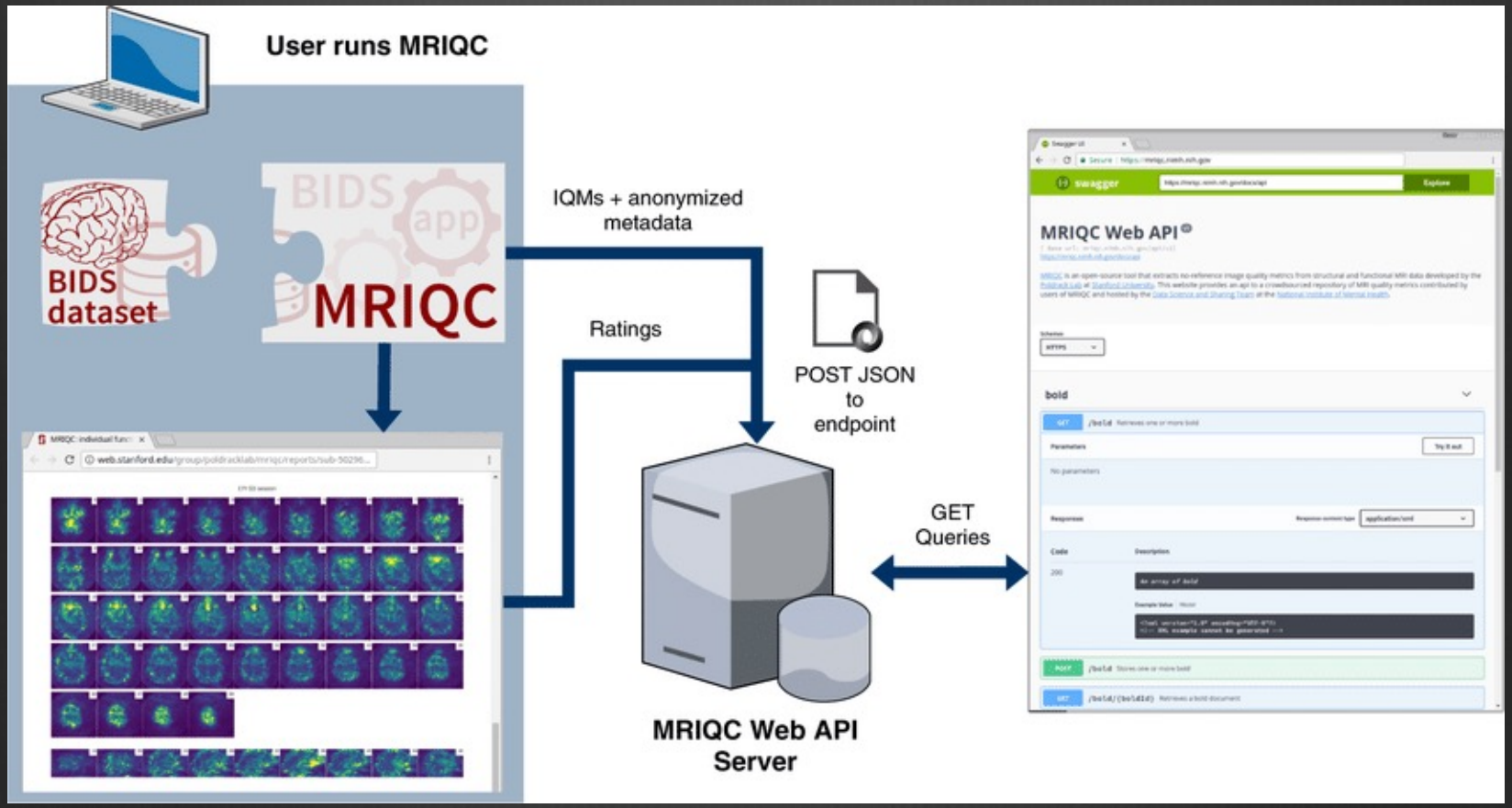
Many Open Science Rubrics require judgement

- FAIR (Wilkinson et al., 2016)
- Delphi Consensus (Cobey et al., 2023)

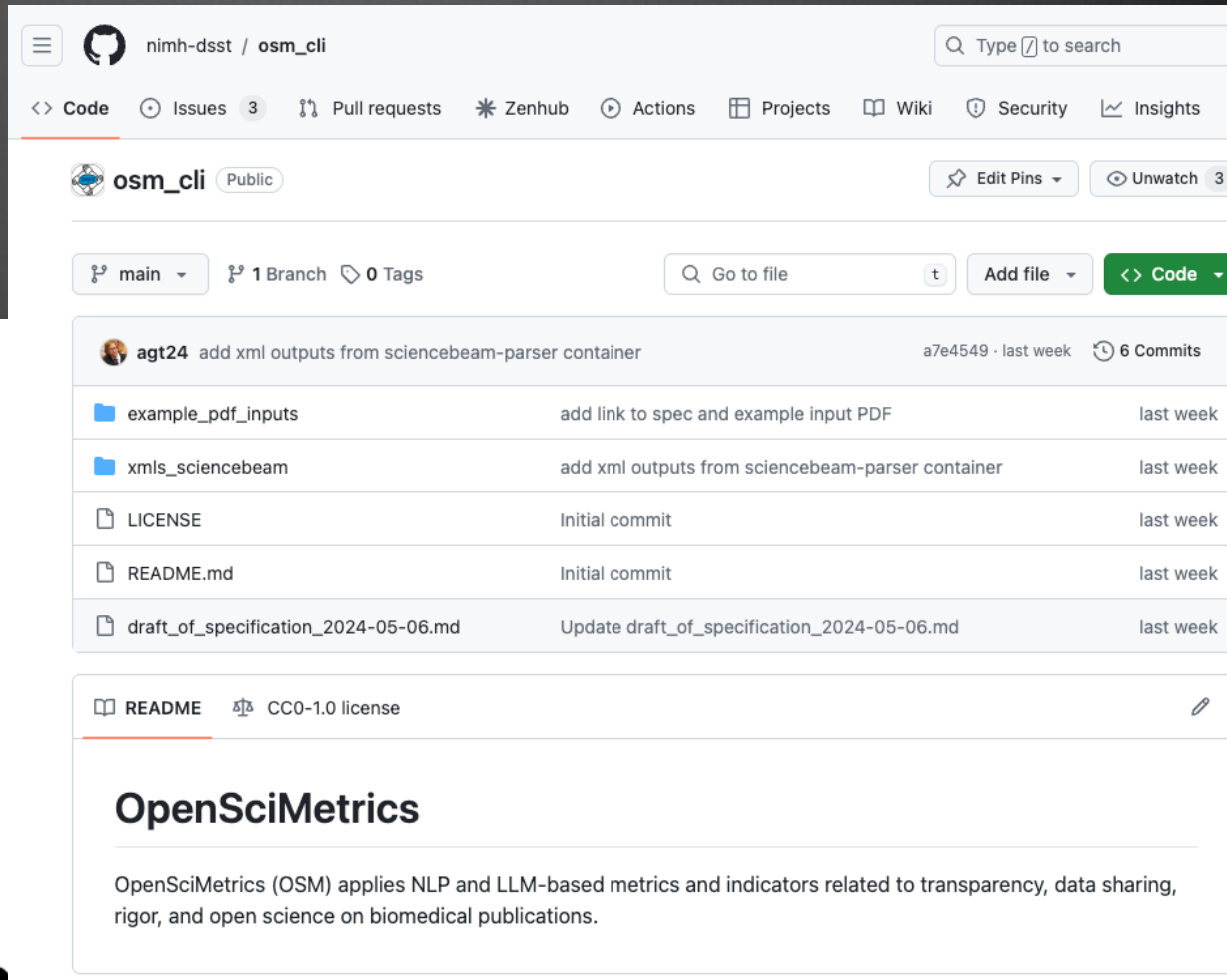
Rank	Practice	Unit of analysis
1	whether clinical trials were registered before they started recruitment	Paper, protocol
2	whether study data were shared openly at the time of publication	paper
3	published open access (time delay?)	paper
4	whether study code was shared openly at the time of publication	paper
5	whether systematic reviews have been registered.	paper
6	were registered clinical trials were reported in the registry within 1 year of study completion	paper
7	whether there was a statement about study materials sharing with publications	paper
8	whether study reporting guideline checklists were used	paper
9	citations to data.	dataset
10	trial results in a manuscript-style publication (peer reviewed or preprint).	paper
11	the number of preprints.	Institution
12	systematic review results in a manuscript-style publication (peer reviewed or preprint).	Systematic review

---

A tool modelled on previous success



Open Science Metrics DB



GitStart

Product of the day
1st

Code as a Service

Assign tickets, get ***high-quality production code*** powered by AI agents and our developer community.

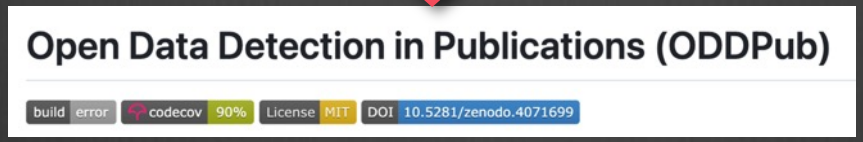
Improving Open Science

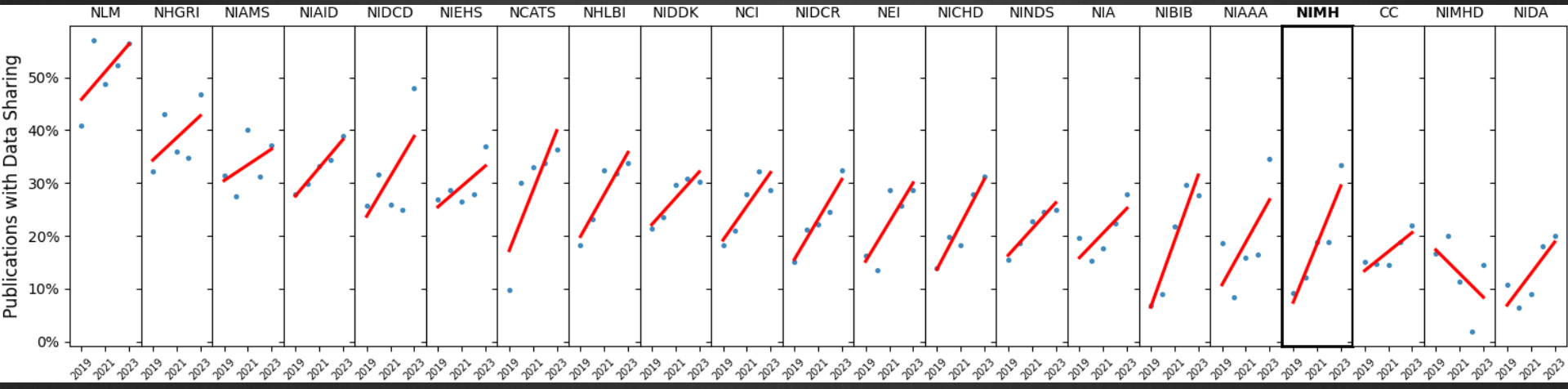
- LLM-based metrics
  - Improved accuracy
  - Extracting more specific details from pubs
  - More nuanced evaluations
- Make metrics uniformly available and more accessible: Open Science Metrics DB

Measuring Data Sharing

- To reward data sharing there must be a mechanism to measure it
- Text mining tools exist to identify data sharing reported in publications
- However, aggregate information on reported data sharing is not easily accessible to funders and policy makers

~25,000 Unique Pubs
2019-2023





Take Homes

- There are many aspects to rigorous, open science
- It can be challenging to keep track
- You should use existing tools to evaluate openness and rigor in every project you're involved in
- These tools are advancing rapidly and play a pivotal role in advancing open science

Slides & links: http://bit.ly/adamt_osrp

Thank You!

Faculty

Jean-Baptiste Poline
Principal Investigator

Researchers

Brent McPherson
Postdoctoral researcher

Nikhil Bhagwat
Academic Associate

Staff

Sebastian Urchs
Research Software Developer

Alyssa Dai
Research Software Developer

Arman Jahanpour
Research Software Developer

Remi Gau
Research Associate

Graduate Students

Kendra Oudyk
PhD student

Ting Zhang
PhD student

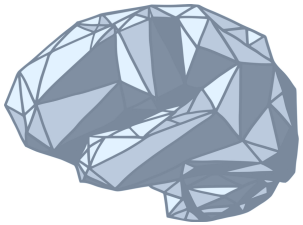
Jacob Sanz-Robinson
PhD student

Mohammad Torabi
PhD student

Emad Askarinejad
PhD student

Niusha Mirhakimi
MSc student

Michelle Wang
PhD student



ORIGAMI
Lab

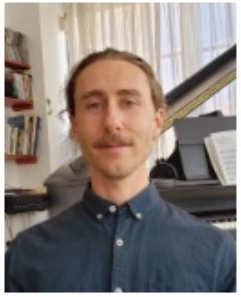
Thank You!

Tanenbaum Open Science Institute (TOSI)

Guy Rouleau
Director, The Neuro
Co-Founder of the Tanenbaum Open Science Institute

Annabel Seyller
Chief of Staff, The Neuro
CEO, Tanenbaum Open Science Institute

Jean-Baptiste Poline



Gabriel Pelletier
Open Science Data Manager
Interim Open Science Alliance Officer

Luisa Pimentel
Open Science Community Officer

NEURO SERIES
KILLAM SEMINARS

Lauréats KILLAM Laureates · neuro

https://openscienceofficehours.github.io/osoh_website/

Take Homes

- There are many aspects to rigorous, open science
- It can be challenging to keep track
- You should use existing tools to evaluate openness and rigor in every project you're involved in
- These tools are advancing rapidly and play a pivotal role in advancing open science

Slides & links: http://bit.ly/adamt_osrp